Evaluating the User Interface of a Ubiquitous Computing system Doorman

Kaj Mäkelä, Esa-Pekka Salonen, Markku Turunen, Jaakko Hakulinen, and Roope Raisamo
Tampere Unit for Computer-Human Interaction (TAUCHI)
FIN-33014 University of Tampere, Finland
+358 3 215 8558
kaj@cs.uta.fi

1. Introduction

We conducted a Wizard of Oz experiment for a speech-based ubiquitous computing system called Doorman ('Ovimies' in Finnish) [4] as a part of our iterative development process. We evaluated our current multimodal and spoken-language interaction models on users before constructing the actual speech recogniser.

Ubiquitous computing systems cannot necessarily be tested in laboratories. Therefore, in many cases the testing of the interfaces has to be done in the actual scene of the action with real life problems. To gather reliable information about human-computer communication it is important to observe the human behaviour in a situation in which they believe to be interacting with a real computer system. [3].

Wizard of Oz testing [1, 3] is an experimental user interface evaluation method in which the user of the system is made to believe that he or she is interacting with a fully implemented system though the whole or a part of the interaction of the system is controlled by a human, a wizard, or several of them. Wizard of Oz tests are useful in supporting design process and evaluating the interface [1, 2, 8]. The method has been commonly used to test natural language dialogue systems [3] and multimodal systems [5, 8]. Here we apply this method to ubiquitous computing applications.

2. System description

The Doorman system is used to help the staff members and visitors in TAUCHI [6] premises in their communicational tasks and everyday lives. This is done by automatically opening the door after recognising the person or the target of the visit, by guiding the visitors to their destination and by conveying organizational and personal messages, such as e-mail messages, to the staff members. The aim of the system is to serve all the users in some way, at least by calling for external help when a problem occurs. The Doorman system has some resemblance to Office Monitor by Yankelovich and McLain [9].

The Doorman uses spoken language as the main modality to communicate with the users. Speech recognition is used as an input and speech synthesis as an output method. The target of the visitor's visit or the identity of the staff member is recognised from their speech.

The system is installed in two locations, one in the front door and one in the lobby. It gathers information about the situation at the front door with a microphone, a doorbell switch and a door micro-switch. The output of the system is presented to the user with synthesised speech via a speaker. Inside, in the lobby the guidance is given in a multimodal way by using speech synthesis together with an anthropomorphic robot pointing to the direction the user should go to find the target. The robot itself has been implemented with three servo-motors controlled by a micro-controller.

The Doorman system is based on a distributed software architecture called Jaspis [7]. Jaspis is a Java-based adaptive speech user interface architecture that has been developed in TAUCHI originally for spoken dialogue applications, but has been expanded to include features that support developing ubiquitous computing applications.

3. Description of the experiment

The aim of the study was to test and analyse the multimodal dialogue and spoken-language model designed for the system before constructing the actual speech recognisers. We wanted to find out how the users actually behave and what kind of language they actually use when talking to this kind of computing system. We were also interested to know how our current guidance to the target of the visit was perceived and understood.

The experiment was conducted by replacing the forthcoming speech recognition module of the Doorman system with a wizard application used manually by the human wizard observing the situation. The speech of the users was recorded and all the system tasks and sensor inputs were logged to be analysed later.

We implemented the application for the wizard to give speech recognition information manually to the system. The control application was designed to be as simple to use as possible to ensure short response times and to minimise the possibility for errors. The tool provides a simple list-based user interface consisting of all the possible alternatives for the speech recognition results. To keep the behaviour of the system consistent and credible we formed a set of rules for the human wizards operating the system.

The test was conducted in five days, one of which was used for training and pilot testing the setup. The test was run approximately 4 hours per day, on a quite varying basis. The test sessions lasted from 45 minutes to 1.5 hours each time. The test was conducted by two persons: one was acting as a wizard and one was gathering permissions from visitors for recording.

4. Results

During the experiment, the system was used in 74 occasions, of which 22 were visitors, and 52 were staff members. The results show that the system prompt was formed so that in most of the visitor cases (77 percents, 17 persons) the users acted and

answered in the way they were expected, and thus the system was able to successively serve these visitors. However, the system failed to serve the visitors in few cases. These visitors either did not spoke at all or searched for a person who was not a member of the group. Visitors were given a possibility to bypass the system and some visitors used it.

The system was prepared to serve two kinds of users: the staff members and the visitors. The staff members and some of the visitors, for example students, have a key or a key card to the premises. Therefore the key holders had a possibility to use other methods to enter the premises and they often used it despite of the request to use the system.

It was shown that in most of the cases the users will choose the easiest and quickest way to handle the task. If the system is not able to serve them rapid enough they will choose an alternative method. For example, long system prompts and delays in the system response were found irritating and the users holding a key or a key card often chose to use one instead of waiting the system to react.

The visitors were assumed to come to TAUCHI to meet someone or to find some room in TAUCHI premises. In the system prompt the users were informed that they should state a name of a person or a room. During the experiment the visitors did not use names of the rooms when stating the target of their visit.

We wanted the system to be easy to use as compared to the use of a key when getting inside. In the dialogue model it was assumed that the staff members are not willing to make any extra effort in order to get inside, for example, to push the doorbell button. However, in 19 percent of the cases the staff members did use the doorbell when starting a dialogue with the system. They therefore heard the prompt designed for the needs of the visitors.

The guide robot was often passed without listening to its instructions. Also the guidance given by the guide robot was found too long, slow and unclear.

The problems we came up with were mainly due to too long system prompts and wrong assumptions of the behaviour of the user. It was also shown that the appearance of the guide robot and the guidance messages need some changes.

We will address the problems occurred during the experiment with the following operations. The system prompts are going to shortened and made clearer. The delays are going to be shortened by rearranging the speech and the action. A feedback is going to be given after a speech input. The error handling is going to be prepared for typical errors and the error loop is going to be shortened to avoid frustration. Also the interaction model is going to be reformed to match the found behaviour of the users. The guidance given by the guide robot is going to be made landmark based and the position and the appearance of the guide robot are going to be altered for better notability.

The users holding a key are accustomed to use their key in order to get inside and therefore do not use the system so often. To gain more data on the use of the system, the system initiative should be increased to make the users to use it.

Two persons acting as wizards managed to keep the operation of the system consistent and correct. The biggest problem was to rapidly decide during the experiment how to handle unexpected speech inputs.

5. Conclusion

The experiment gave us information on the actual use and problems occurred during the operation of the system, which are valuable when developing the interface further. However, the experimental data was quite limited and further testing should be conducted before drawing generalising conclusions. The experiment showed us that the Wizard of Oz method is effective in evaluating the interface of a Ubicomp system.

The most important findings that help in the further development of the system were related to the structure of the system prompts, the need for system initiative and better error handling, and the need for changes in the way the guidance is arranged with the robot.

Based on our experience we recommend using Wizard of Oz method during the iterative development of ubiquitous computing system user interface. However, the process is very time-consuming and it demands good planning and training.

6. References

[1] Bernsen, N.O., Dybkjær, H., and Dybkjær, L., *Designing Interactive Speech Systems*. Springer-Verlag, London. 1999, 127-160.

[2] Bretan, I., Ereback, A.-L., MacDermid, C., and Waern, A., Simulation-based dialogue design for speech-controlled telephone services. *CHI'95 Interactive Posters*, available at http://www.acm.org/sigs/sigchi/ chi95/proceedings/intpost/ib_bdy.htm.

[3] Dahlbäck, N., Jönsson, A., and Ahrenberg, L., Wizard of Oz Studies – Why and How. *Proceedings of the 1993 International Workshop on Intelligent User Interfaces (IUI'93)*, ACM Press, 1993, 193-200.

[4] The Doorman system http://www.cs.uta.fi/hci/spi/Ovimies/index_en.html.

[5] Salber, D., and Coutaz, J., Applying the Wizard of Oz Technique to the Study of Multimodal Systems. Proceedings of the EWHCI'93, Third International Conference. *Lecture Notes in Computer Science*, Vol. 753, Springer-Verlag, 1993, 219-230.

[6] Tampere Unit for Computer-Human Interaction http://www.cs.uta.fi/hci/.

[7] Turunen, M., and Hakulinen, J., Jaspis - A Framework for Multilingual Adaptive Speech Applications. CD ROM *Proceedings of 6th International Conference of Spoken Language Processing (ICSLP 2000)*, Beijing, China, 2000.

[8] Wyard, P., and Churcher, G., A realistic Wizard of Oz simulation of a multimodal spoken language system. *Proceedings of 5th International Conference on Spoken Language Processing (ICSLP'98)*, Sydney, Australia, 1998.

[9] Yankelovich, N., and McLain, C.D., Office Monitor. *CHI 1996 Conference Companion*, ACM Press, 1996, 173-174.